An Inconsistent Bayes Estimator in Bivariate Survival Curve Analysis

By

Ronald C. Pruitt

Technical Report No. 529

School of Statistics

University of Minnesota

24 July 1989

[1]Research partially supported by National Institutes of General Medical Sciences Grant GM-42921-01.

Abstract

An example of an inconsistent sequence of Bayes estimates is presented. This example occurs in a very natural framework when estimating a bivariate survival function from incompletely observed data. The same example shows that the generalized maximum likelihood estimator in this problem is not unique even for large sample sizes and may not have a limit. For this example there does exist a particular sequence of generalized maximum likelihood estimates which does converge to the true underlying distribution.

Suppose we are interested in the distribution of a bivariate random vector (T_1, T_2), but we are unable to make observations from this random vector directly due to a nuisance censoring variable (C_1, C_2) which is independent of (T_1, T_2). Instead we are able to observe X_1, X_2, D_1, and D_2, where $X_j = \min[T_j, C_j]$ and $D_j = 1[X_j = T_j]$. We discuss the behavior of a nonparametric Bayesian estimator of the distribution of (T_1, T_2) based on independent and identically distributed sample data $(X_{1i}, X_{2i}, D_{1i}, D_{2i})$ for $i = 1, \ldots, n$.

If we place a Dirichlet process prior on (T_1, T_2) the posterior distribution is a mixture of Dirichlet processes (Pruitt, 1988). Here we pursue an example in which the sequence of Bayes estimates obtained under squared error loss is inconsistent for the distribution of (T_1, T_2). This is another example where Bayes estimators can be inconsistent. Examples of such behavior have been pursued by Freedman (1963), Freedman and Diaconis (1983), Doss (1985a,b), and Diaconis and Freedman (1986a,b). This example shows the Dirichlet process prior can lead to inconsistent results in a very natural framework with incomplete data. The modelling assumptions seem reasonable as does the Bayesian analysis. There is no other known method of Bayesian analysis of this problem which gives consistent results.

Here is a specific example. Let (T_1, T_2) be uniformly distributed over $[1,2] \times [1,2] \cup [2,3] \times [2,3]$, and let (C_1, C_2) have mass 1/3 at (4,4), 1/3 at (0,4), and 1/3 at (4,0). Assume the prior measure of the Dirichlet process is $M\alpha_0$ where α_0 is a probability measure. Let α_0 be uniform over $[0, c] \times [0, c]$ for some c. Let P_T denote the distribution of (T_1, T_2) and let P_U denote the uniform distribution over $[1,3] \times [1,3]$. We will show the Bayes estimator converges to $1/3P_T + 2/3P_U$.

Let $\theta_i = (T_{1i}, T_{2i})$. For each observation $(X_{1i}, X_{2i}, D_{1i}, D_{2i})$ there is an associated set A_i in which θ_i is constrained to occur. The sets A_i have one of the following four forms:

$$
\begin{array}{llll}
\text{I.} & A_i = & \{c_1\} \times \{c_2\} & \text{if } D_{1i} = D_{2i} = 1 \\
\text{II.} & A_i = & [c_1, \infty) \times \{c_2\} & \text{if } D_{1i} = 0 \text{ and } D_{2i} = 1 \\
\text{III.} & A_i = & \{c_1\} \times [c_2, \infty) & \text{if } D_{1i} = 1 \text{ and } D_{2i} = 0 \\
\text{IV.} & A_i = & [c_1, \infty) \times [c_2, \infty) & \text{if } D_{1i} = D_{2i} = 0.
\end{array}
$$

The posterior distribution is a mixture of Dirichlet processes with parameter measures $M\alpha_0 + \sum_i \delta_{\theta_i}$ and mixing distribution the conditional distribution of $(\theta_1, \ldots, \theta_n)$ given

$\theta_1 \in A_1, \ldots, \theta_n \in A_n$. The Bayes estimator with respect to squared error loss is

$$\frac{M}{M+n}\alpha_0 + \frac{n}{M+n}\mathrm{E}[n^{-1}\sum \delta_{\theta_i} | \theta_1 \in A_1, \ldots, \theta_n \in A_n].$$

See Pruitt (1988). The limit as $n \to \infty$, if it exists, is the same as the limit of the conditional expected value, which we now examine.

Renumber so that $\theta_1, \ldots, \theta_{R_{1n}}$ are the observations of type I, $\theta_{R_{1n}+1}, \ldots, \theta_{R_{1n}+R_{2n}}$ are the observations of type II, and the remainder of the observations are of type III. Note (R_{1n}, R_{2n}) has a multinomial distribution with parameters n and (1/3,1/3). Let $R_{3n} = n - R_{1n} - R_{2n}$. The mass assigned from the uncensored observations is

$$n^{-1}\sum_{i=1}^{R_{1n}} \delta_{\theta_i},$$

which converges to 1/3 P_T since these θ_i are independent, identically distributed from (T_1, T_2) and R_{1n} is binomial with parameters n and 1/3.

We now turn to the mass assigned to the points of types II and III. With probability one, none of the uncensored points intersect any of the sets A_i from the observations of type II or III. These sets A_i form a grid of rays with each set of type II intersecting each set of type III. There are $R_{2n}R_{3n}$ intersection points, and from symmetry considerations each will receive the same amount of mass, say I_n (intersection). The mass assigned to each of the sets A_i of type II which is not assigned to the intersection points will also be the same, say A_n (across). Let D_n (down) be the common amount of mass assigned to each of the sets A_i of type III which is not assigned to the intersection points. We wish to show that the amount of mass assigned to the non-intersection points becomes negligible as $n \to \infty$. For this it suffices to assume $R_{3n} \leq R_{2n}$ and show

$$\frac{R_{2n}A_n + R_{3n}D_n}{R_{2n}R_{3n}I_n} \leq \frac{2D_n}{R_{3n}I_n} \doteq \frac{6D_n}{nI_n} \to 0 \tag{1}$$

as $n \to \infty$. To do this we need expressions for D_n and I_n. Let $E_{rcn} = A_{R_{1n}+r} \cap A_{R_{1n}+R_{2n}+c}$ and let $F_{cn} = A_{R_{1n}+R_{2n}+c} \setminus \cup_i E_{icn}$. The sets E_{rcn} are the intersection points and the sets F_{cn} are the remainder of the sets of type III which are not intersection points. Note that

$$D_n = n^{-1}\,\Pr[\theta_{R_{1n}+R_{2n}+1} \in F_{1n} | \theta_1 \in A_1, \ldots, \theta_n \in A_n],$$

and

$$(2) \qquad I_n = 2n^{-1}\Pr[\theta_{R_{1n}+1} \in E_{11n}, \theta_{R_{1n}+R_{2n}+1} \in E_{11n} | \theta_1 \in A_1, \ldots, \theta_n \in A_n].$$

We first need to develop some facts about conditional probabilities from the Dirichlet process, and the results about D_n and I_n will then follow from counting arguments. If we try and rewrite (2) using unconditional probabilities we run into problems since the probabilities involved are zero. A regular conditional distribution can be found using the methods of Pfanzagl (1979) as given in Pruitt (1988). Heuristically the result can be understood by approximating the sets A_i of types II and III by sets of width ϵ, A_i^ϵ, which have positive α_0 measure. If we do this,

$$\Pr[\theta_{R_{1n}+i} = \theta_{R_{1n}+R_{2n}+i}, i = 1, \ldots, k, \theta_i \in A_i^\epsilon, i = 1, \ldots, n] \doteq \frac{(\beta\epsilon^2)^{R_{1n}}(\beta\epsilon^2)^k 1^k (\gamma\epsilon)^{R_{2n}+R_{3n}-2k}}{M^{(n)}}$$

where $\beta = c^{-2}$ is the height of the density of α_0 at any intersection point, $\gamma = c^{-1}$ is the integral of the density over any of the sets A_i (or F_{cn}) of type II or III, and $M^{(n)} = M(M+1)\cdots(M+n-1)$. Note that the conditional probability of observing exactly k intersection points does not depend on which intersection points are observed due to symmetry. The limiting argument goes through as the heuristics indicate to give

$$\Pr[\theta_{R_{1n}+i} = \theta_{R_{1n}+R_{2n}+i}, i = 1, \ldots, k | \theta_1 \in A_1, \ldots, \theta_n \in A_n] \propto \beta^k \gamma^{-2k}$$

where the proportionality constant does not depend on k.

For our example note that $\beta^k\gamma^{-2k} = 1$. Recall that we are assuming $R_{2n} \geq R_{3n}$ and write

$$I_n \propto 2 \sum_{k=1}^{R_{3n}} l_k$$

where l_k is the number of ways to get k intersections with one being specified. Here

$$l_k = \binom{R_{3n}-1}{k-1} \frac{(R_{2n}-1)!}{(R_{2n}-k+1)!}.$$

Similarly

$$D_n \propto \sum_{k=0}^{R_{3n}-1} m_k = \sum_{k=0}^{R_{3n}-1} \binom{R_{3n}-1}{k} \frac{R_{2n}!}{(R_{2n}-k+1)!}$$

where m_k is the number of ways to get k intersections with a specific type III observation being specified as unequal to any of the type II observations.

Using Stirling's formula,

$$\lim_{n\to\infty} \frac{\sum_{k=\lceil \lambda R_{3n}\rceil}^{R_{3n}} l_k}{\sum_{k=1}^{R_{3n}} l_k} = 1$$

for any $0 < \lambda < 1$, since terms away from R_{3n} decrease in magnitude exponentially. Also note that for $k \geq \lambda R_{3n}$, $m_k \leq \lambda^{-1}(1-\lambda)R_{2n}l_k$ which gives

$$\frac{\sum_{k=\lceil \lambda R_{3n}\rceil}^{R_{3n}} m_k}{2n\sum_{k=\lceil \lambda R_{3n}\rceil}^{R_{3n}} l_k} \leq \frac{(1-\lambda)R_{2n}}{2\lambda n} \to \frac{(1-\lambda)}{6\lambda}.$$

This is enough to show (1), since $\liminf 6D_n/nI_n \geq 0$ and $\limsup 6D_n/nI_n \leq (1-\lambda)/\lambda$ for any $0 < \lambda < 1$. This shows that the mass assigned to the nonintersection points of the grid goes to zero. Thus the weight assigned to the points of type II and III converges to a uniform distribution over $[1,3] \times [1,3]$ since the intersection points are uniformly distributed over this region. The Bayes estimate converges to $1/3P_T + 2/3P_U$.

This inconsistency occurs because the sets of type II and III do not gain any information about the distribution of (T_1, T_2) through the sets of type I because of the absolute continuity of the distribution. A solution which is currently being explored is the use of smoothing to correct this problem. The same kind of result holds for the generalized maximum likelihood estimator (GMLE) studied by Muñoz (1980), Campbell (1981), and Hanley and Parnes (1983). In this example the GMLE is not uniquely defined and has no limit. Any distribution which assigns mass $1/n$ to each set of type I, mass $1/R_{2n}$ to each set of type II, and mass $1/R_{3n}$ to each set of type III is a GMLE. All GMLE's assign mass $1/n$ to each of the uncensored observations. For instance the distribution assigning mass $1/R_{2n}R_{3n}$ to each of the intersection points is a GMLE.

There is a sequence of GMLE's which does converge to the correct limit. To see this, consider the distribution which assigns mass $2/R_{2n}R_{3n}$ each to only half of the intersection points. We only consider the case when R_{2n} and R_{3n} are each even for simplicity. We choose half of the intersection points by finding a and b such that half of the sets of type II have second coordinate less than b and half of the sets of type III have first coordinate less than a. Then half the intersection points are contained in $[1,a] \times [1,b] \cup [a,3] \times [b,3]$. This is a GMLE. As $n \to \infty$, a and b both converge to 2 and the GMLE converges to the true distribution.

References

CAMPBELL, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68** 417-422.

DIACONIS, P., and FREEDMAN, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1-26.

DIACONIS, P., and FREEDMAN, D. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68-87.

DOSS, H. (1985a). Bayesian nonparametric estimation of the median; part I: Computation of the estimates. *Ann. Statist.* **13** 1432-1444.

DOSS, H. (1985b). Bayesian nonparametric estimation of the median; part II: Asymptotic properties of the estimates. *Ann. Statist.* **13** 1445-1464.

FREEDMAN, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math Statist.* **34** 1386-1403.

FREEDMAN, D., and DIACONIS, P. (1983). On inconsistent Bayes estimates in the discrete case. *Ann. Statist.* **11** 1109-1118.

HANLEY, J.A., and PARNES, M.N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39** 129-139.

MUÑOZ, A. (1980). Nonparametric estimation from censored bivariate observations. *Technical Report* **60**, Stanford University.

PRUITT, R.C. (1988). Bayesian survival curve estimation for incompletely observed multivariate data. *Technical Report* **516**, University of Minnesota.

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF MINNESOTA
206 CHURCH ST. S.E.
MINNEAPOLIS, MN 55455